

FFCUL/LASIGE TEAM SECOND YEAR REPORT

The Epiwork project started in February 2009 and will run for 48 months. According to the work plan FFCUL is involved in the following work packages:

- WP 1 Population Models and Contact Networks
- WP 3 Information platform
- WP 4 Epidemic Modelling Platform
- WP 7 Management

The FFCUL participation in Epiwork involves two groups:

- CMAF Group
- LASIGE Group.

This report summarises the progress in the third semester of the project activity by the LASIGE Group, which leads WP3 and participates in WP4 and WP7.

The LASIGE Team working on WP3 and W4 in this period includes:

- Mário J. Silva (FCUL Faculty, worked in the reporting period in WP3, WP4 and WP7)
- Fabrício Silva (FCUL Faculty, worked in the reporting period in WP3 and WP4 – Left FFCUL at the end of July 2010)
- Francisco Couto (FCUL Faculty, worked in the reporting period in WP3)
- Dulce Domingos (FCUL Faculty, worked in the reporting period in WP3 – started in May/2010)
- Luís Lopes (Post-doc researcher, worked full-time during the reporting period until August, 2010 and part-time since September, 2010; he left the project at the end of October).
- Hugo Ferreira (Graduate student, worked full-time in the reporting period until the end of August/2010).
- Carla Patrícia Sousa (Graduate student, worked full-time in the reporting period).
- João Zamite (Graduate student, worked full-time in the reporting period).
- Juliana Duque (Graduate student, started full-time collaboration in September, 2010).
- João Ferreira (Graduate student, started half-time collaboration in November, 2010).

There is a discrepancy in the profiles of the people hired to work on Epiwork at LASIGE and the initially planned, the reason being that we were unable to fill the software engineer and a post-doc positions as planned, with previous directly related experience and salaries matching our budget.

As a result, we started using Informatics Engineering and Biomedical Informatics students on the infrastructure setup assisted by FCUL faculty and senior technical staff. LASIGE has hired more people in the first two years at a reduced cost, which was the contingency measure found to manage the situation.

Meanwhile, Fabrício Silva and Hugo Ferreira announced that they would leave the project at the end of July 2010, and Luís Lopes at the end of October 2010.

Prof. Dulce Domingues joined the project in the reporting period. She will be mainly addressing data privacy issues in the epidemic marketplace, together with João Zamite, who will be advised on is PhD by her.

We have hired a new Master student in Communication Design, Juliana Duque, who will work in the project for one year in the design of the user interfaces for a second version of the Epidemic Marketplace. Juliana started collaborating in EPIWORK in September.

João Ferreira, a new PhD student in Informatics, has been working half-time for Epiwork (paid by LASIGE internal funds) since November 2010. His research is on ontologies and he will be working on interlinking epidemic data and contributing to the work to be reported on the Epiwork deliverable D3.5 Epidemic data ontology, due January 31st 2012.

1. WP3 — Information platform: Work progress, collaborations and achievements during the period

This Work Package is lead by FFCUL, with a total contribution of 82 persons.month (60 hired technicians + (22) academics).

2.1 Background Information on WP3

Work package number	3	Start date or starting event:				Month 1		
Work package title	Information platform							
Activity type	RTD							
Participant number	12	1	2	3	4	5	9	10
Participant short name	FFCUL	ISI	FGC-IGC	TAU	MPI-DS	AIBV	BIU	FBK-IRST
Person-months per participant	82	60	4	8	42	2	11	6

The whole WP3 activity is structured into four tasks:

Task 3.1 – Data Collection.

Participants: FFCUL, ISI, FBK-IRST, BIU, MPI-DS, FGC-IGC, AIBV.

Description: Realistic simulations of epidemic processes crucially depend on the availability of datasets describing human behaviour and pathogen-host interactions. Datasets include population movement data, social and behavioural data, health related data, geographic data, detailed geo-temporal epidemic incidence and immunization data, pathogen evolution and multi-strains circulation data. Data can come from a variety of different sources, including hospital records, country statistics, Web content, and others. It can range from a global scale, such as the worldwide air transportation infrastructure, down to the detailed description of individual **activities** at a minute-by-minute scale. This task will create a catalogue of databases of epidemiological data across Europe, with extensive meta-data describing the main characteristics of the available information sources. This catalogue

will be integrated with a collaborative platform that will be set up for online discussion and exchange of meta-data among the participants.

Task 3.2 – Meta-Model Design.

Participants: FFCUL, IGC, ISI.

Description: While some of the previously mentioned datasets are freely available on the Web (e.g. WHO Global Health Atlas, Eurostat), they are often scattered in different repositories, cover partial regions of the world and come in different formats, according to different standards and classifications. The project envisions a unified and integrated approach for the management of these resources, with the design and implementation of an Epidemic Marketplace Platform, publicly available on the web. The platform supports the sharing and management of epidemic datasets and resources as well as their rating, annotation, and selection. It is an on-line social networking site that will serve researchers, practitioners, and educators all over the world to foster a virtual community for epidemic research. It will support the exchange of resources as well as user interactions. Based on a Web2.0 approach, users will become active participants, sharing information and data, and collaborating online, rather than being satisfied with a passive information consumer/viewer role. We envision proposing a simple reference format, which will facilitate the navigation and use of the datasets. Each dataset will come with a metadata file, signalling general metadata for resource management, containing data such as: the title, the date of submission, version, the source of the data and coverage. Moreover, the metadata will include information for a more thorough description of the data included in the dataset, providing a framework for a more specific description, for example, of epidemiologic and geographic data,. The Marketplace will support flexible and intuitive tools for navigation and selection of resources. Standard classifications as well as tagging systems proposed by users will be supported.

Task 3.3 – Epidemic Marketplace Platform.

Participants: FFCUL, ISI.

Description: This task will implement a platform based on the integration of grid technology and publicly available services and software on the web to support the sharing and management of epidemic datasets and resources as well as their rating, annotation, and selection. The Epidemic Marketplace Platform will be an on-line social networking site that will serve researchers, practitioners, and educators all over the world to foster a virtual community for epidemic research. It will support the exchange of resources as well as user interactions. Based on some of the Web2.0 characteristics, users will become active participants, generating information and providing data for sharing, and collaborating online,

rather than being satisfied with a passive information consumer/viewer role. More specifically, researchers can use and contribute to the Marketplace in several different ways. They can: (1) use it as a catalogue of data sources containing the metadata describing existing databases; (2) view, download, tag, and comment on the available resources; (3) provide compliant datasets and relevant information; (4) use it as a forum where to publish information about their own data, seek modellers to collaborate with, share and distribute their new findings.

Task 3.4 – Evaluation and monitoring of the use of the catalogue and collaboration services.

Participant: FFCUL.

Description: This task involves the monitoring of epidemiological data exchanges performed through the mediating services platform. The evaluation will assess not only the coverage of the catalogued resources, but the users' satisfaction with the user interface and integrated collaborative tools made available through the epidemiological marketplace platform. More importantly, the analysis of the collected datasets and their annotations and usage will provide a rich environment for deriving an epidemiology ontology, which will help further on the integration and communication among the community of epidemiologists.

2.2 Progress in the Reporting Period

- *A summary of progress towards objectives, and details for each task the team was involved;*
- *Highlight clearly significant results;*
- *Team publications within the scope of the project (please provide a pdf of the publication if possible);*
- *List of outreach activities (conferences, Invited talks, presentations, workshops, tutorials,);*
- *List of press releases or media coverage, any particular dissemination activity;*
- *unanticipated finding, opportunity etc*

Main Activities at FFCUL in the second year of the project in WP3:

1. Mário Silva and Fabrício Silva attended the Epiwork Review in Brussels in March, 2010
– Mário Silva presented the two completed WP3 deliverables to the EC and gave an overview of the progress and challenges in WP3. The demo of the fully functional Epidemic Marketplace prototype was presented to the reviewers, who had the opportunity to obtain user accounts for browsing the EM repository and forum.
2. **[outreach]** Mário Silva and Fabrício Silva attended the COSI-ICT cluster workshop in Brussels, March, 2010.
3. **[opportunity]** Francisco Couto participated in the VPH-OBO ontology workshop at Cambridge UK for knowing relevant case studies and guidelines in ontology development in the biomedical field. During the workshop, Francisco discussed with Barry Smith (National Center for Ontological Research, USA) the state of the art of epidemiological ontologies and a possible collaboration within the IDO (infectious disease ontology).
4. **[outreach]** Mário Silva attended WP5 Meeting in Amsterdam in May; presentation by Mário Silva and discussions on the Integration and privacy issues regarding collected data on WP3/WP5.

5. Major reorganization of the base software infrastructure of the Epidemic Marketplace, which is now entirely deployed in various virtual machines hosted in the two dedicated servers plus and two network storage units.
6. Metadata model refinement, identification of relevant ontologies for annotation and development of a metadata editor for the repository embodying the proposed model. The metadata model is fully implemented on the metadata editor that is now part of the production version of the Epidemic Marketplace. The implementation of the policies outlined in deliverable D3.1 has been completed.
7. Development of a second version of the data collector, MedCollector, which is running on the production version of the Epidemic Marketplace.
8. Development of the initial mediator services, based on Muradora. There were some issues related to the format of service requests and results, regarding the incompatibilities of two main requirements: adoption of the REST syntax and OAI-PMH. The web services are now available in the production version of the Epidemic marketplace.
9. Evaluation of Integration of Drupal with Fedora Commons, using the Islandora module extension to Drupal and using this software as basis for a new version of the Repository/Forum of the Epidemic Marketplace.
10. **[outreach]** Mário J. Silva attended the ASSYST Perada Workshop - Towards a Science of Socially Intelligent ICT in London, August 2010, where he made the case for handling privacy from the ground-up in information systems collecting and managing social networking data (such as Epiwork).
11. **[outreach]** Mário J. Silva, Dulce Domingos, Patrícia Sousa and João Zamite attended ECCS'10, the European Conference on Complex Systems in Lisbon, September 2010, where they participated in the COSI-ICT workshop.
12. WP3/WP4 meeting in Lisbon on September 16 at ECCS'10. The contents of the EM were discussed with Daniela Paolotti, who submitted the data files used by the GLEaMviz simulator, with focus on appropriate metadata for the categorization of the resources. Some concepts behind the development of EM v2.0 were discussed, with focus the user-oriented interface. There was some brainstorming on the integration between WP3 and WP4 was held. An initial approach to integration was discussed which consisted of a Client Application that checks the EM for updates to the subscribed collections and downloads them to the WP4 server. GLEaMviz would then run

simulations using the data and the Client will upload the Flash Visualization Files back to the Epidemic Marketplace. This was the idea behind the SimpleEMClient that we developed later.

13. **[significant result]** Public Release of the Epidemic Marketplace (Deliverable 3.3) at <http://epimarketplace.net>. Anyone from the community can now define collections and retrieve/upload datasets.
14. Mário Silva, Dulce Domingos and João Zamite attended the Second Epiwork Meeting in Torino, December 2010. Mário Silva presented the status of WP3 and Zamite introduced Privacy + Access Control and the plans of handling the issue at the Epidemic Marketplace.
15. In a WP3/WP4 meeting in Torino December the WP3 and WP4 members discussed plans for further integration of the data and computational platforms. In a new scenario to be demonstrated using the EM platform and the GLEaMviz simulator, simulation files will be deposited in the EM, then the sent to the simulation platform and finally the results will transferred from the computation platform to the Epidemic Marketplace. The simulation parameters and the results would be shared by the users and computational tools having appropriate access permissions.
16. **[opportunity]** Initiated design of new access control method to be implemented in the EM. EM users will be able to specify user access permissions based on their Social Networks membership. This has enormous potential for simplifying the process of managing access permissions. These ideas were initially presented at the 2nd Epiwork Meeting in Torino, in December 2010.
17. **[publication]** In January, Zamite et al. submitted an extended version of the paper “MEDCollector: Multisource Epidemic Data Collector” to a special issue of the Springer journal LNCS Transactions on Large-Scale Data and Knowledge-Centered Systems entitled Database Systems for Bio-medical Applications.
18. SimpleEMClient available for download from the Epidemic Marketplace at the end of January, 2010.
19. EM 2.0 platform development underway for release in 2011. Major features: new user interface and component integration based on the Drupal Content-Management System, improved access control management, redesigned web-services.

Publications and Presentations:

1. **[publication]** Mário J. Silva, Francisco M. Couto, Dulce Domingos, Juliana Duque, Hugo Ferreira, Luís F. Lopes, Daniela Paolotti, Fabrício Silva, Patrícia Sousa, João Zamite, D 3.3 Public Release of the Epidemic Marketplace Platform Technical Report. Technical Report. LASIGE, University of Lisbon, Faculty of Sciences, September 2010. Also available from the epiwork.edu website.
2. **[publication]** Luis Filipe Lopes, A Metadata Model for the Annotation of Epidemiological Data Master Thesis, University of Lisbon, Faculty of Sciences, September 2010.
3. **[publication]** João Zamite, Multisource Epidemic Data Collector Master Thesis, University of Lisbon, Faculty of Sciences, September 2010.
4. **[publication]** Hugo Ferreira, O Mediador do Epidemic Marketplace Master Thesis, University of Lisbon, Faculty of Sciences, January 2011.
5. **[publication]** Fabrício A.B. Silva, Mário J. Silva, Francisco Couto 2010: Epidemic Marketplace: an e-Science Platform for Epidemic Modelling and Analysis. ERCIM News 82(), 43-44. Special Theme: Computational Biology.
6. **[publication]** Fabrício A.B. Silva, Mário J. Silva, Francisco Couto 2010: Epidemic Marketplace: an e-Science Platform for Epidemic Modelling and Analysis. ERCIM News 82 - Special Theme: Computational Biology.
7. **[publication + presentation]** Luis Filipe Lopes, Fabrício A.B. Silva, Francisco Couto, João Zamite, Hugo Ferreira, Carla Sousa, Mário J. Silva, Epidemic Marketplace: An Information Management System for Epidemiological Data. Presented at ITBAM'10 - 1st International Conference on Information Technology in Bio- and Medical Informatics - DEXA 2010 - August, 2010.
8. **[publication + presentation]** João Zamite, Fabrício A.B. Silva, Francisco Couto, Mário J. Silva, MEDCollector: Multisource Epidemic Data Collector. Presented at ITBAM'10 - 1st International Conference on Information Technology in Bio- and Medical Informatics - DEXA 2010 - August, 2010.
9. **[publication + presentation]** Mário J. Silva, Fabrício A.B. Silva, Luís Filipe Lopes, Francisco Couto, Building a Digital Library for Epidemic Modelling. Proceedings of ICDL 2010 - The International Conference on Digital Libraries 1, p. 447--459, New

Delhi, India, 23--27 February, 2010. TERI Press -- New Delhi, India. Presentation of invited paper.,

10. **[presentation]** Presentation and demo of EPIWORK to students at the University of Valencia, Spain in an invited 4hrs seminar, by Fabricio Silva, June 2010.
11. **[presentation]** Presentation of EPIWORK to students of the Master/Phd in Epidemiology at the Faculty of Medicine of the University of Lisbon, in an invited 3h seminar, by Mário J. Silva.
12. **[presentation]** Mário J. Silva, Privacy in Socially Intelligent ICT. Imperial College, London, UK. ASSYST Perada Workshop - Towards a Science of Socially Intelligent ICT. August, 2010.

Activities at FFCUL in the second year of the project in Task 3.1:

The data collection activity started at the end of the first year. We have been automatically collecting the following data with the MedCollector since then, from the following sources:

- Messages from Twitter containing disease names and posted at each country or capital of the world.
- Google Flu Trends Datasets.
- RSS Feeds from the CDC. Including travel notices alerts and emerging diseases alerts.
- Email Newsletter messages from ProMEDmail.

These datasets are now being periodically assembled and uploaded into the Epidemic marketplace.

Activities at FFCUL in the second year of the project in Task 3.2:

Work in this task included:

- Population of the repository with datasets and resources found relevant to epidemic modellers gathered from the Web, used for validation of the metadata modelling decisions.
- Development of a metadata editor for epidemic datasets for the EM platform, following the policies outlined in deliverable D3.1.

Activities at FFCUL in the second year of the project in Task 3.3:

- Reconfiguration of the infrastructure, now based on virtual machines. Configuration of the production and development environments for the Lisbon EM node.
- Deployment of a new version of the EM base software, to be based in Fedora Commons version 3.0 and the Drupal Content Management System, initiated in January continued. This new version will substitute the current, based on Fedora Commons 2.2.2 and Muradora 1.3.3. This new version is planned to replace the production environment of the EM in 2011.
- The first operational prototype of the Epidemic Marketplace was presented in the March 2010 review in Brussels.
- Discussions with partners involved in WP3 on how to identify relevant datasets to the catalogue and strategies and incentives for populating the Epidemic Marketplace.
- WP3/WP4 collaboration on the integration of the computational platform with the data platform.
- Public release of the Epidemic Marketplace at <http://epimarketplace.net>. Anyone from the community can now define collections and retrieve/upload datasets.
- Initiated design of new access control system to be implemented in the EM.
- Development of the SimpleEMClient, available for download from the Epidemic Marketplace.
- EM 2.0 platform development underway for release in 2011. Major features: new user interface and component integration based on the Drupal Content-Management System, improved access control management, redesigned web-services.

Activities at FFCUL in second of the project in Task 3.4:

- Continued work on planning of the monitoring and log data collection and analysis tasks. Collection of usage data is now active since the release of the Epidemic Marketplace to the public. This is done through Google analytics and analysis of the web servers log data.

2.3 Effort Allocation

The effort allocated by LASIGE to WP3 in the reporting period is as follows:

WP3	FFCUL	Effort (p.m)
Reporting Period		Non-Perm + Perm =Total
Year 1: February 1, 2009 to January 31, 2010		19.38 + 4.53 = 23.91
Year 2: February 1, 2010 to January 31, 2011		32.55 + 6.43 = 38.98

The commitment to the project in the first year was 23.91 persons.month, 19.38 pm from technicians, and 4.53 p.m from permanent staff.

In the reporting period (2nd year), the commitment was 38.98 persons.month, 32.55 pm from technicians, and 6.43 pm from permanent staff

The effort allocated is now well above planned (assuming a 10.25 p.m constant effort/semester), and reflects the decision of hiring more (and less skilled) staff for the FFCUL/LASIGE team of Epiwork in Semester 2.

2. Work in WP4 — Epidemic Modelling Platform

This task is lead by ISI, with a total FFCUL contribution of 19 persons.month (12 hired + (7) academics).

Activities at FFCUL in the third semester of the project in WP4:

The LASIGE work on WP4 is related to the integration of the Computational and Data Plaforms of Epiwork, which has been accounted in detail in the description of activities of LASIGE on WP3. Below, we mention the most significant of these activities:

- Mário Silva and Fabrício Silva participated in the Project Review in Brussels, March 2010.
- Mário Silva and Fabrício Silva attended the COSI-ICT cluster workshop in Brussels, March 2010.
- Discussions in Amsterdam, May 2010, at the WP5 meeting;
- Discussions at the Complex Systems Conference, Lisbon, September 2010.
- Discussions at the ASSYST Perada Workshop - Towards a Science of Socially Intelligent ICT in London, August 2010,
- Discussions in Torino at the 2nd Episwork meeting, December 2010.
- Validation of the meta-data uploaded by ISI, discussions on how EM data would be accessed from the computational platform tools and support in the process.
- Development of the SimpleEMClient for use by WP4 tools.
- Initiated design of scenario for running GLEaMViz simulations through the Epidemic Marketplace.

2.4 Effort Allocation

The effort allocated to WP4 in the reporting period is as follows:

WP4	FFCUL	Effort (p.m)
Reporting Period		Perm + Non-perm =Total

WP4	FFCUL	Effort (p.m)
Reporting Period		Perm + Non-perm =Total
Year 1: February 1, 2009 to January 31, 2010		5.72+0.50 = 6.22
Year 2: February 1, 2010 to January 31, 2011		5.05 + 0.40 = 5.45

The commitment to the project in the first year was 6.22 persons.month, 5.72 pm from technicians, and 0.50 p.m from permanent staff.

The commitment to the project in the second year (the reporting period) was 5.45 persons.month, 5.05 pm from technicians, and 0.40 p.from permanent staff.

This is as planned, reflecting the fact that this task is led by ISI and our contribution will intensify later, as we settle on a common software architecture, including interfaces, for orchestrating and running services on both platforms.

The reported effort by technicians reflects the splitting of the effort dedicated to setting-up the hardware and base software of the epiwork infrastrucure in Lisbon between WP3 and WP4 in the first year, and, in the second year, the development of the SimpleEMClient for synchronizing local folders with EM streams and initial designs for integration of GleamViz with the Epidemic Marketplace.

Work in WP7 — Management

This task is lead by ISI.

FFCUL Effort in this task: 4 persons.month.

Activities at FFCUL in the second year of the project in WP7:

1. Data collection activities for preparation of the 3rd semester and 2nd year reports, and preparation of management presentations.
2. Mário Silva and Fabricio Silva attended the Epiwork Review in Brussels, March 2010.
3. Mário Silva attended the 2nd Epiwork Meeting in Torino, December 2010.

2.5 Effort Allocation

The effort allocated to WP7 in the reporting period is as follows:

WP7 Reporting Period	FFCUL	Effort (p.m) Perm + Non-perm =Total
Year 1: February 1, 2009 to January 31, 2010		1.10 + 0.0 = 1.10
Year 2: February 1, 2010 to January 31, 2011		0.96 + 0.0 = 0.96

The effort dedicated to the project in the period, 0.96 p.m., was 100% contributed by permanent staff, as planned.